# Cyclic-Coded Integer-Forcing Equalization

Or Ordentlich and Uri Erez, *Member, IEEE*

*Abstract*—A discrete-time intersymbol interference channel with additive Gaussian noise is considered, where only the receiver has knowledge of the channel impulse response. An approach for combining decision-feedback equalization with channel coding is proposed, where decoding precedes the removal of intersymbol interference. This is accomplished by combining the recently proposed integer-forcing equalization approach with cyclic block codes. The channel impulse response is linearly equalized to an integer-valued response. This is then utilized by leveraging the property that a cyclic code is closed under (cyclic) integer-valued convolution. Explicit bounds on the performance of the proposed scheme are also derived.

*Index Terms*—Linear Gaussian channels, single-carrier modulation, intersymbol interference, combined equalization and coding, cyclic codes, decision-feedback equalization.

## I. INTRODUCTION

The intersymbol interference (ISI) channel with additive Gaussian noise is one of the most basic channel models arising in digital communications. Thus, considerable effort has been devoted to developing effective transmission schemes for this channel; see, e.g., [1], for a comprehensive survey. The channel is described by

$$y_k = x_k + \sum_{m \neq 0} h_m x_{k-m} + n_k$$
$$= x_k + \mathrm{ISI}_k + n_k, \tag{1}$$

where $\mathrm{ISI}_k$ is the ISI resulting from other data symbols, and $n_k$ is additive white Gaussian noise (AWGN) with zero mean and unit power.

The channel model may further be characterized by the availability of channel state information (CSI), where we distinguish between the case where CSI is available to the transmitter and the receiver alike and the case where CSI is available to the receiver only. As we next briefly recall, while the distinction between these two cases does not make a great difference at high signal-to-noise ratios (SNR) (which is the main focus of this paper) in terms of capacity [2] (i.e., whether water-filling may be performed or not), it is of significant consequence for the design and implementation of equalization and coding schemes.

In the past decades coding for AWGN channels has reached an advanced state, and practical coding schemes (e.g., turbo and LDPC codes) operating near capacity are known. It is thus desirable to combine AWGN coding and decoding

techniques with equalization in a *modular* way, with the aim of approaching the capacity of the ISI channel.

The multitude of approaches developed to achieve reliable communication over the ISI channel may be roughly divided into two classes: multi-carrier approaches and single-carrier approaches. Both approaches may in principle be used to approach the capacity of the ISI channel, but offer different practical tradeoffs as we briefly touch upon next.

In multi-carrier transmission, the ISI channel is transformed into a set of parallel AWGN subchannels, each subchannel corresponding to a different frequency bin and experiencing a different SNR. This approach has the advantage that the subchannels are (virtually) ISI free, and thus the problems of equalization and decoding are decoupled. However, it has some drawbacks: the alphabet size of the transmitted symbols is considerably enlarged, which in turn makes the approach inapplicable to some media, such as magnetic recording channels. A related phenomena associated with multi-carrier transmission is that it results in a high peak-to-average power ratio which may also be undesirable (see e.g., [3], [4]). Furthermore, when CSI is available only at the receiver, bit allocation is precluded, and channel coding and decoding become more difficult, due to the variation of the SNR across subchannels.

Single-carrier approaches try to eliminate most of the ISI without severely increasing noise power. The simplest approach is linear equalization consisting only of a "feed-forward" equalizer (FFE), which roughly transforms the channel into an additive colored Gaussian noise channel, where the minimum mean-squared error (MMSE) criterion corresponds to (linearly) maximizing the signal-to-interference-plus-noise (SINR) at the "slicer". Performance may be improved using (non-linear) decision-feedback equalization, in addition to FFE. Specifically, MMSE decision-feedback equalization (DFE) will be discussed in greater detail in the next section. In fact, as shown by Guess and Varanasi [5], the MMSE-DFE architecture is optimal in the sense of attaining mutual information, and allows to approach capacity with AWGN encoding/decoding, if decisions (fed to the DFE) are based on codewords rather than symbols. See also [6], [7].

Unfortunately, the Guess-Varanasi approach, while quite pleasing from a theoretical perspective, requires long inter-leaving as well as long zero-padding, which in turn requires long latency. This drawback can be avoided if CSI is available at the transmitter by Tomlinson-Harashima precoding ( [2], [8]), which essentially moves the DFE to the transmitter, but is inapplicable if the transmitter has no knowledge of the channel.

The approach proposed in this paper allows to avoid error propagation *without* incorporating an interleaver and with CSI available at the receiver only. In essence, the proposed

Or Ordentlich and Uri Erez are with the Department of Electrical Engineering-Systems, Tel Aviv University, Ramat Aviv 69978, Israel. E-mail: {ordent,uri}@eng.tau.ac.il.

method enables to perform (soft or hard) block decoding *before* decision feedback is performed.

We build on the integer-forcing (IF) equalization approach, which was recently proposed [9] in the context of general MIMO channels with CSI available at receiver only. In this approach, multiple streams are encoded using an identical linear code and the receiver equalizes the channel matrix to any full-rank integer matrix rather than to the identity.

In the context of IF equalization, an ISI channel may be viewed as a Toeplitz matrix. This special structure allows to further replace the multiple codewords (in the context of ISI, this number would be large) with a *single* codeword provided that the linear code is further a *cyclic code*.[1]

The paper is organized as follows. Section II recalls some basic results on single-carrier equalization. Section III describes IF equalization as well as the general structure of the proposed scheme. Section IV derives the criteria for choosing the FFE equalizer. Section V derives bounds on the attained performance. In Section VI, the performance of the scheme is analyzed, and some examples are given. Section VII discusses practical coding techniques for integer-forcing equalization at high transmission rates. The paper concludes with Section VIII.

## II. Preliminaries

We briefly review basic single-carrier equalization architectures. In the sequel, we use $D$-transform notation for sequences; i.e., a sequence $\{s_k\}$ is represented by its $D$-transform $S(D) = \sum_k s_k D^k$. For example the channel (1) may be expressed as,

$$Y(D) = H(D)X(D) + N(D).$$

The simplest criterion for linear equalization (LE) is that of Zero-Forcing (ZF), where the ISI is completely canceled using an FFE only. This corresponds to taking the front end (linear) filter to be

$$A_{\text{ZF}}(D) = \frac{1}{H(D)},$$

resulting in the equalized channel response $G(D) = 1$. The induced noise enhancement can be large, especially when $H(D)$ has zeros near the unit circle. A variant that takes into account both ISI and noise enhancement is the linear MMSE equalizer

$$A_{\text{MMSE-LE}}(D) = \frac{H^*(D^{-*})}{H(D)H^*(D^{-*}) + 1/\text{SNR}}.$$

The MMSE-LE suffers from smaller (and in particular bounded) noise enhancement while allowing some residual ISI. The MMSE criterion is equivalent to maximizing the SINR at the slicer input [6].

Decision-feedback equalization (DFE) (see Figure 1) is based on using previously detected symbols in order to cancel the induced ISI from the symbol entering the slicer. In this approach, if all previous data symbols are detected without

error, then postcursor ISI can be removed. Specifically, the output of the FFE in Figure 1 is given by

$$
\begin{aligned}
y'_k &= x_k * g_k + z_k \\
&= x_k + \sum_{m=-\infty}^{-1} x_{k-m}g_m + \sum_{m=1}^{\infty} x_{k-m}g_m + z_k \\
&= x_k + \text{ISI}_k^{\text{pre}} + \text{ISI}_k^{\text{post}} + z_k,
\end{aligned}
$$

where $G(D)$ is the equivalent channel after the operation of the FFE. The DFE then subtracts $\hat{\text{ISI}}_k^{\text{post}} = \sum_{m=1}^{\infty} \hat{x}_{k-m}g_m$ from $y'_k$, where $\hat{x}_k$ are decisions on past transmitted symbols, giving rise (assuming correct past decisions) to the equivalent channel

$$y''_k = x_k + \sum_{m=-1}^{-\infty} x_{k-m}g_m + z_k.$$

MMSE-DFE is optimal, when using the optimal FFE, in the sense that the SINR at the slicer (assuming correct past decisions) is equal to the SNR of an AWGN channel having the same capacity as that of the ISI channel [6]. Combining DFE with coding, however, is a non-trivial task. Since a decision on the value of the last symbol $\hat{x}_k$ must enter the feedback loop at every time instance, there is an intrinsic tension with the latency required for channel coding. Many approaches have been suggested in order to overcome this obstacle (see for example [10]–[14]), but to the best of our knowledge none of them allow to exchange the order of decoding and ISI removal which is the aim of the present work. Doing so, *directly* addresses the basic problem of ensuring that reliable decisions enter the DFE loop.

## III. Combining Cyclic Codes With IF Equalization

We begin this section by recalling IF equalization in the context of ISI channels. For a more general description, see [9]. For purposes of simplicity of exposition, we limit ourselves in the sequel to real-valued channels (and transmission of real symbols). The extension of the proposed scheme to complex transmission is straightforward, and, unless stated otherwise, all results derived in this work hold for the complex case as well. Further, we first describe the ZF-IF approach which is suitable for the high SNR regime, and relegate the extension to MMSE-IF (which is suitable for any SNR) to the Appendix.

An IF equalizer (depicted in Figure 2), rather than attempting to cancel the ISI (as in linear ZF as well as ZF decision feedback), ensures that the ISI is restricted to take only integer values. Assuming the data symbols are taken from a constellation of size $q$ consisting of the integers $\{0, 1, \ldots, q-1\}$ (i.e., a PAM constellation), which we may identify as the ring $\mathbb{Z}_q$, the ISI will take only integer values provided that the channel impulse response is equalized to an impulse response $I(D)$, such that all the coefficients of $I(D)$ are integers. We restrict attention to responses $I(D)$ of finite length and further assume that $I(D)$ is a monic[2] polynomial

---

[1]The crucial element needed is that it be a linear shift invariant code, not necessarily cyclic.

[2]There is some loss of generality in this assumption. However, the optimal choice for $I(D)$ (as we discuss in Section IV) is almost always monic. When this isn't the case, the proposed approach is not very effective. The restriction that $I(D)$ be monic can be removed with some modifications to the scheme.
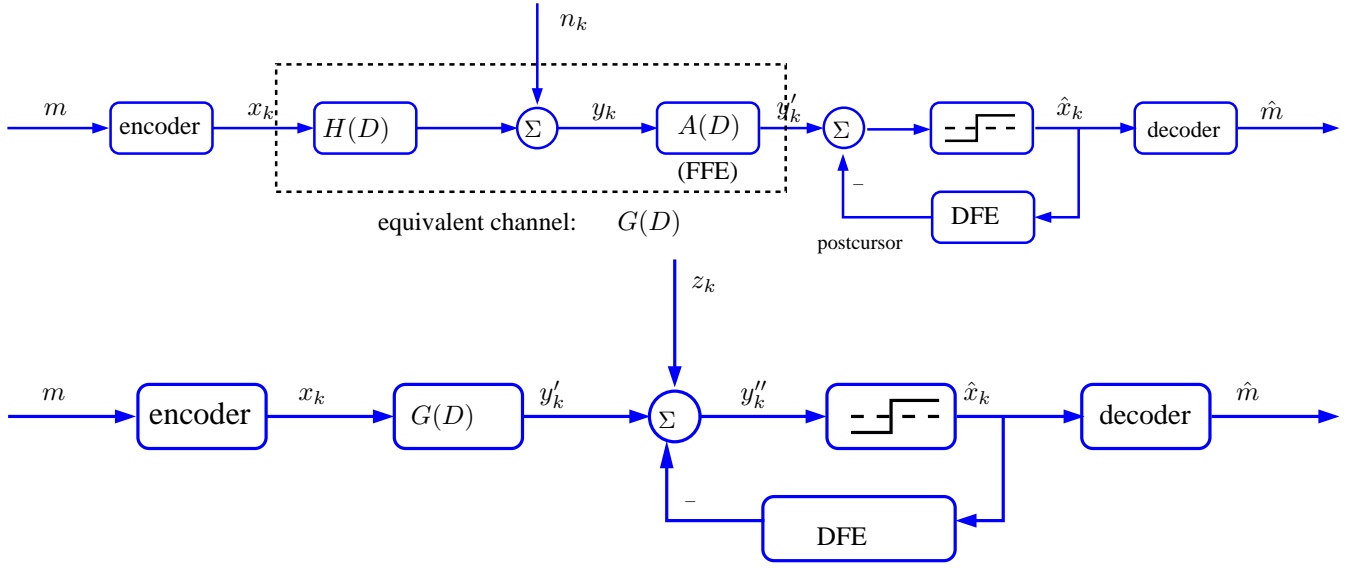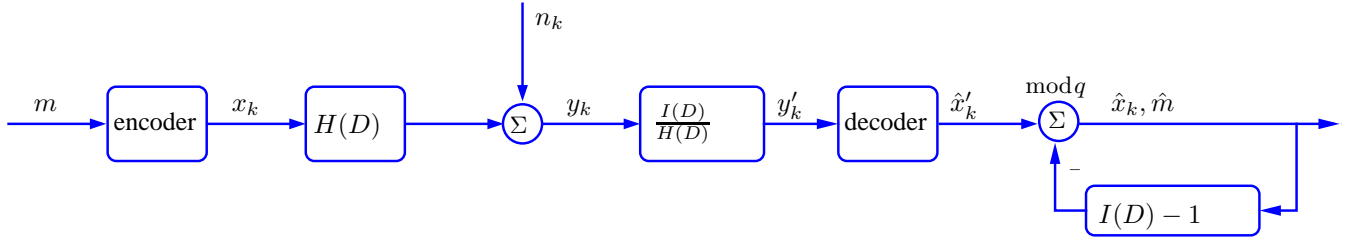
Fig. 1. Decision-feedback equalization.



Fig. 2. A schematic description of an IF-DFE system.

(with integer coefficients), i.e.,

$$I(D) = \sum_{k=0}^{n-1} i_k D^k,$$

where $i_0 = 1$. We denote the vector of coefficients of $I(D)$ by,

$$\mathbf{i} = [1 \ i_1 \ \ldots \ i_{n-1}].$$

Note that taking $I(D) = 1$ is a special case in which the IF equalizer reduces to a ZF-LE one. In most cases, choosing $I(D)$ otherwise results in smaller noise enhancement. The criterion for determining the feed-forward filter for the IF equalizer will be given in the next section. We further note that IF equalization is closely related to partial response signaling (see for example [15]), and was previously suggested in [16] where a suboptimal algorithm for finding a good integer channel $I(D)$ was also suggested. Another work of similar spirit is [17] where it is shown that in some cases the complexity of the maximum likelihood decoder that is matched both to the channel and the code can be reduced if the channel is integer-valued.

The key feature of IF equalization exploited in this paper is that, as we observe below, if the transmitted data is taken from a *cyclic* code, then the output of the equalized channel is also a member of the codebook.

Assume some monic integer-valued filter was chosen by the receiver (the considerations for choosing $I(D)$ are discussed

in Section IV). The FFE applied as the front-end filter is therefore,

$$A_{\text{ZF-IF}}(D) = \frac{I(D)}{H(D)}. \tag{2}$$

The output of the equalizer is

$$Y'(D) = X(D)I(D) + Z(D),$$

where $Z(D) = N(D)I(D)/H(D)$ is colored Gaussian noise. We further define the SNR to be,

$$\mathsf{SNR}_{\text{ZF-IF-DFE}} = \frac{\sigma_x^2}{\sigma_z^2} = \frac{\sigma_x^2}{\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{|I(e^{jw})|^2}{|H(e^{jw})|^2}d\omega}.$$

The impulse response $I(D)$ should be chosen so as to maximize $\mathsf{SNR}_{\text{ZF-IF-DFE}}$. The definition of SNR (i.e, which does not include the residual ISI as noise) is justified when using a cyclic code as described next.

*Definition 1:* A linear block code $\mathcal{C}$ of length $N$ over $\mathbb{Z}_q$ is called cyclic, if for every codeword $\mathbf{x} \in \mathcal{C}$, all cyclic shifts of $\mathbf{x}$ are also codewords in $\mathcal{C}$.

We denote cyclic convolution w.r.t. block length $N$ by $\otimes$. The following proposition is an immediate consequence of Definition 1.

*Proposition 1:* Let $\mathcal{C}$ be a cyclic code of length $N$ over $\mathbb{Z}_q$. Then for any vector $\mathbf{i}$ of length $N$ with integer entries,

$$\mathcal{C} \otimes \mathbf{i} \subseteq \mathcal{C}.$$

That is, $\mathcal{C}$ is closed under integer-valued cyclic convolution over $\mathbb{Z}_q$.

We conclude that if the linear convolution $x_k * i_k$ over $\mathbb{R}$ were a cyclic convolution $x_k \otimes i_k$ with operations over $\mathbb{Z}_q$, then we would be able to decode $\mathbf{x}' = \mathbf{x} \otimes \mathbf{i}$ directly from the output $\mathbf{y}'$, and then reconstruct $\mathbf{x}$ from it.[3]

The transformation to cyclic convolution may be accomplished in much the same way as in DMT/OFDM transmission. In order to do so, we use a systematic cyclic encoder at the transmitter such that the last $K$ symbols of a $[N, K]$ (cyclic) block code are the data symbols and the first $N - K$ symbols are redundancy symbols. We fix the last $n - 1$ information symbols to be zeros, where $n$ is the length of $I(D)$. Consequently, *all* codewords end with $n - 1$ zeros. The effective rate is reduced to $(K - n + 1)/N$.

Due to the zero padding, if we take only the first $N$ samples of the output of the FFE, reduced modulo $q$, we get,

$$y'_k \bmod q = [(\mathbf{x} \otimes \mathbf{i})_k + z_k] \bmod q, \quad \text{for } k = 1, \ldots, N. \quad (3)$$

This process is illustrated in Figure 3. The output of the FFE is processed in blocks of length $N$ (the blocks play no role in the operation of the FFE itself). The zero padding, in addition to initializing the DFE with zeros, as described below, ensures that when a new data block is transmitted, the "channel's memory" is empty, i.e., there is no ISI between consecutive blocks.
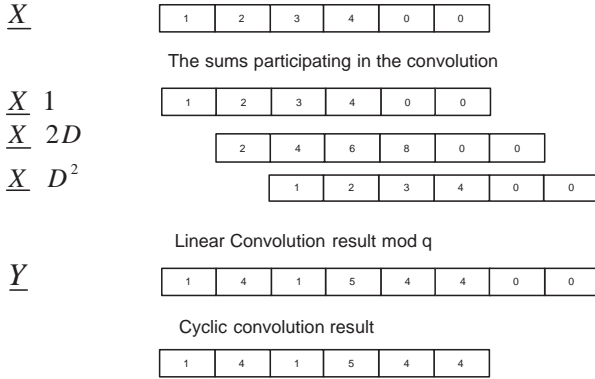


Fig. 3. An Illustration of transforming a linear convolution $x_k * i_k$ over $\mathbb{R}$ into a cyclic convolution $x_k \otimes i_k$ over $\mathbb{Z}_7$. In this example $I(D) = 1 + 2D + D^2$.

Since the result of the cyclic convolution is itself a codeword, we may apply the channel decoder directly to it. Assuming correct decoding, we next reconstruct $x_k$ from the decoded result of the cyclic convolution. This can be done in a recursive way by applying a DFE as shown in Figure 2.[4] Since we know the last $n - 1$ values of $x_k$ (i.e., they are all zero), we are able to recover $x_k$ for every $k = 1, ..., N - (n - 1)$.

---

[3] It is important to note that in order to reconstruct $\mathbf{x}$ from $\mathbf{x} \otimes \mathbf{i} \bmod q$, all elements of $\mathbf{x}$ must be in $\mathbb{Z}_q$. Nevertheless, the elements of $\mathbf{i}$ need only be integers but need not necessarily be restricted to $\mathbb{Z}_q$.

[4] All operations of the DFE are carried over $\mathbb{Z}_q$ It is always possible to reconstruct $x_k$ regardless of $q$ (i.e. whether it is prime or not) and the choice of the integer-valued filter since $I(D)$ is monic.

The reconstruction of $x_k$ from $x_k \otimes i_k$ is performed using a DFE rather than a linear equalizer, as the integer filter $I(D)$ often has zeros on the unit circle, and hence it is not invertible.

We end this section by remarking that the recently proposed "Signal Codes" [18], though not cyclic, are also suitable for IF equalization. Signal codes are a special family of lattice codes which are generated by convolving sequences of data symbols taken from a PAM/QAM constellation with a monic filter $F(D)$. For data sequences of length $N$ and an encoding filter $F(D)$ of length $M$, the generating matrix of the lattice code is an $(N + M - 1) \times N$ Toeplitz matrix. For a data sequence $B(D)$ the corresponding codeword is[5] $X(D) = B(D)F(D)$. When IF equalization is used in conjunction with signal codes, the output of the equivalent channel after the FFE is (neglecting the additive noise) $X'(D) = B(D)I(D)F(D)$. Let $B'(D) = B(D)I(D)$, and note that $B'(D)$ is a sequence of $(N + n - 1)$ PAM/QAM symbols as $I(D)$ is an integer valued filter of length $n$. It follows that $X'(D)$ is a codeword from a signal code generated by $F(D)$ with a generating Toeplitz matrix of dimensions $(N + M + n - 2) \times (N + n - 1)$. For large enough $N$, this codebook has similar performance to those of the original codebook. Therefore $B'(D)$ can be decoded and then $B(D)$ can be reconstructed from it. Note that for signal codes there is no need to transform the linear convolution performed by the channel into a cyclic one, as done when cyclic codes are used. In [18] it is demonstrated that signal codes can operate reasonably close to capacity with an acceptable complexity, and may therefore be attractive for IF equalization at high SNR.

## IV. CRITERIA FOR CHOOSING THE FILTER $I(D)$

We wish to find an integer-valued filter $I(D)$ of length $n$ such that the noise enhancement experienced by the ZF-IF equalizer for a given channel $H(D)$ is minimized. The noise variance at the output of the IF front-end filter (2) is,

$$\sigma^2_{\text{ZF-IF-DFE}} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{|I(e^{j\omega})|^2}{|H(e^{j\omega})|^2} d\omega. \quad (4)$$

Denote,

$$K(D) = \frac{1}{H(D)H^*(D^{-*})}. \quad (5)$$

Thus, $k_m$ is an autocorrelation sequence (and in particular $k_{-m} = k_m$).

Using straightforward algebra, (4) may be written as a quadratic form,

$$\sigma^2_{\text{ZF-IF-DFE}} = \mathbf{i}\tilde{K}_n\mathbf{i}^{\mathbf{T}},$$

where $\tilde{K}_n$ is the positive semi-definite Toeplitz matrix,

$$\tilde{K}_n = \left[ \begin{array}{cccc} k_0 & k_{-1} & k_{-2} \cdots & k_{-(n-1)} \\ k_1 & k_0 & k_{-1} \cdots & k_{-(n-2)} \\ \vdots & \vdots & \vdots \ddots & \\ k_{n-1} & k_{n-2} & k_{n-3} \cdots & k_0 \end{array} \right].$$

---

[5] In fact, some additional "shaping" steps must be taken in order to assure that the power constraint is preserved. The full details can be found in [18].

Let $F_n$ be any matrix satisfying,

$$\tilde{K}_n = F_n{}^T F_n.$$

We therefore have,

$$\sigma_{\text{ZF-IF-DFE}}^2 = \mathbf{i}\tilde{K}_n \mathbf{i^T} = \mathbf{i}F_n{}^T F_n \mathbf{i^T} = \|F_n \mathbf{i^T}\|^2. \quad (6)$$

Equation (6) implies that finding the optimal (ZF) integer-valued filter $I(D)$ is equivalent to finding the shortest vector in the lattice $\Lambda(F_n)$, which is composed of all integral combinations of the columns of $F_n$, i.e.,

$$\Lambda(F_n) = \{\lambda = F_n \mathbf{i} \mid \mathbf{i} \in \mathbb{Z}^n\}. \quad (7)$$

Finding the shortest lattice vector is known to be NP hard, but fortunately efficient suboptimal algorithms for finding a short lattice basis are known. An important representative of this class of algorithms is the celebrated LLL algorithm [19], which has polynomial complexity and usually gives adequate results in practice. In order to find a "good" integer-valued filter $I(D)$, we may therefore apply the LLL algorithm on $F_n$. The algorithm's result is a new basis (of short vectors) for $F_n$. We then need to find the shortest vector $\mathbf{v}$ in this basis and choose $\mathbf{i} = F_n{}^{-1}\mathbf{v}$. In the next section we derive an upper bound on the induced noise enhancement when the true (optimal) shortest vector of the lattice is used, which serves as a useful benchmark.

## V. AN UPPER BOUND ON THE NOISE ENHANCEMENT OF OPTIMAL ZF-IF EQUALIZATION

We now upper bound the noise enhancement induced by optimal ZF-IF equalization. Throughout this Section we assume that the channel has a finite-length impulse response of length $p + 1$, i.e.,

$$H(D) = \sum_{k=0}^{p} h_k D^k.$$

The length of the channel will play an important role in the derived bound.

We proceed to upper bound $\sigma_{\text{ZF-IF-DFE}}^2$ using known results from the theory of lattices, and the theory of Toeplitz matrices.

*Lemma 1:* (Minkowski) Let $\lambda_1(\Lambda(F_n))$ denote the shortest vector in the $n$-dimensional lattice $\Lambda(F_n)$. If $F_n$ is of rank $n$, then,

$$\lambda_1(\Lambda(F_n)) \leq \left(\frac{2^n}{\beta_n}\right)^{\frac{1}{n}} |\det(F_n)|^{\frac{1}{n}},$$

where $\beta_n$ is the constant of proportionality of the volume of an $n$-dimensional sphere, i.e., the volume of an $n$-dimensional sphere with radius $R$ is given by $V_n(R) = \beta_n R^n$.

*Proof:* See, e.g., [20]. ∎

The constant $\beta_n$ can be bounded by (see, e.g., [21]),

$$\beta_n > \left(\frac{2\pi e}{n}\right)^{n/2} \frac{1}{\sqrt{1.4\pi n}}.$$

Recalling that $\tilde{K}_n = F_n{}^T F_n$, we therefore have,

$$\lambda_1^2(\Lambda(F_n)) < \eta(n) \cdot \left[\det(\tilde{K}_n)\right]^{\frac{1}{n}}, \quad (8)$$

where

$$\eta(n) = \frac{2n}{\pi e} \cdot (1.4\pi n)^{\frac{1}{n}}. \quad (9)$$

Therefore, in order to obtain a closed-form bound on $\lambda_1^2(\Lambda(F_n))$, it suffices to evaluate $\det(\tilde{K}_n)$. To that end, define,

$$\alpha_H = \frac{|z_0 z_1 \dots z_{p-1}|^{2p}}{\prod_{\mu,\nu} |z_\mu^* z_\nu - 1|}, \quad (10)$$

where $z_0, z_1, \dots, z_{p-1}$ are the maximum-phase zeros of $H(D)H^*(D^{-*})$ (i.e., the zeros outside the unit circle).

We are now ready to present Theorem 1 which is the main result of this section. The theorem makes use of a result from the theory of Toeplitz matrices which we state first, and the proof of which can be found in [22, Chapter 5].

*Lemma 2:* Let $H(D) = \sum_{k=0}^{p} h_k D^k$ be a polynomial of degree $p$. Further, let $K(D)$ be as defined in (5). Then for every $n \geq p + 1$,

$$\left[\det(\tilde{K}_n)\right]^{\frac{1}{n}} = (\alpha_H)^{\frac{1}{n}} \cdot \exp\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} \log K(e^{j\omega})d\omega\right],$$

where $\alpha_H$ is defined in (10).

*Theorem 1:* Assume that the channel has a finite-length impulse response $H(D) = \sum_{k=0}^{p} h_k D^k$. For an optimal choice of the integer filter $I(D)$, the noise power at the output of the filter $I(D)/H(D)$ is bounded by[6]

$$\sigma_{\text{ZF-IF-DFE}}^2 \leq \sigma_{\text{ZF-DFE}}^2 \cdot \min_{n \geq p+1} \left[\eta(n) \cdot (\alpha_H)^{\frac{1}{n}}\right], \quad (11)$$

where $\eta(n)$ is defined in (9) and $\alpha_H$ is defined in (10).

*Proof:* Lemma 2 may be rewritten as

$$\left[\det(\tilde{K}_n)\right]^{\frac{1}{n}} = (\alpha_H)^{\frac{1}{n}} \cdot \sigma_{\text{ZF-DFE}}^2, \quad (12)$$

where

$$\sigma_{\text{ZF-DFE}}^2 = \exp\left[-\frac{1}{2\pi}\int_{-\pi}^{\pi} \log|H(e^{j\omega})|^2 d\omega\right] \quad (13)$$

is the variance of the AWGN after the optimal FFE is applied in ZF-DFE (see, e.g., [6]). Substituting (12) into (8) yields:

$$\lambda_1(\Lambda(F_n))^2 \leq \eta(n) \cdot (\alpha_H)^{\frac{1}{n}} \cdot \sigma_{\text{ZF-DFE}}^2.$$

Since $\sigma_{\text{ZF-IF-DFE}}^2 = \lambda_1^2(\Lambda(F_n))$, the theorem is proved. ∎

The additional noise enhancement caused by the ZF-IF equalizer w.r.t. an optimal ZF-DFE as bounded in (11) consists of two factors: $(\alpha_H)^{\frac{1}{n}}$ and $\eta(n)$. The factor $(\alpha_H)^{\frac{1}{n}}$ is greater (or equal) to 1 and tends to 1 as $n \to \infty$. Thus, allowing for a long filter $I(D)$ mitigates the effect of this factor. On the other hand, the factor $\eta(n)$ is (approximately) linearly increasing with $n$. The minimization in (11) strikes a balance (i.e., searches for the optimal tradeoff) between these two factors.

---

[6]For (complex) transmission over a complex channel, (9) changes to:

$$\eta(n) = \frac{4n}{\pi e} \cdot (2.8\pi n)^{\frac{1}{2n}}.$$

The tightness of the bound depends on the tightness of the Minkowski bound of Lemma 1. It is known that there exist lattices that satisfy this bound with equality, but for "most" lattices the shortest lattice vector is much shorter than what this bound predicts. We further note that the family of lattices considered in this paper is not general, as $F_n^T F_n$ is a Toeplitz matrix, and therefore the Minkowski bound may never be tight, as we further discuss in the next section.

## VI. PERFORMANCE OF ZF-IF EQUALIZATION

The capacity of the Gaussian ISI channel (1) at high SNR is given by (see e.g. [6])

$$C = \frac{1}{2} \log_2 \left( \frac{\text{SNR}}{\sigma^2_{\text{ZF-DFE}}} (1 + o(1)) \right), \qquad (14)$$

where $\sigma^2_{\text{ZF-DFE}}$ was defined in (13), and where $o(1) \to 0$ as $\text{SNR} \to \infty$. Note that (14) is valid only for channels for which $\sigma^2_{\text{ZF-DFE}}$ is finite, i.e. for channels satisfying the Paley-Wiener condition (see [2]).

In this section we analyze the total gap-to-capacity of the ZF-IF equalization scheme at high SNR, i.e. the gap between the performance obtained using ZF-IF and the optimal performance (14).

The ZF-IF scheme, as described in Section III, transforms the original Gaussian ISI channel into an equivalent modulo-additive channel

$$Y'(D) = [X(D)I(D) + Z(D)] \bmod q$$
$$= [X'(D) + Z(D)] \bmod q,$$

where $Z(D)$ is filtered Gaussian noise. As we recall, correct decoding of $X'(D)$ ensures correct reconstruction of $X(D)$. In practice, the transmitted signal $x_k$ is subject to the power constraint

$$E\left[x_k^2\right] \leq \text{SNR},$$

and hence, instead of mapping the codewords from the cyclic code $\mathcal{C}$ to the constellation $\mathbb{Z}_q$, we map them to the constellation

$$c\sqrt{\text{SNR}} \left\{ \frac{-(q-1)}{2q}, \frac{-(q-3)}{2q}, \dots, \frac{q-1}{2q} \right\}. \qquad (15)$$

The constant $c$ is chosen such that the power constraint is satisfied with equality for a uniform distribution over the constellation. It can be easily verified that $c > \sqrt{12}$ for any value of $q$ and approaches $\sqrt{12}$ for large values of $q$.[7]

Let $\Delta = c\sqrt{\text{SNR}}$. Using the constellation (15), the equivalent modulo-additive channel after the FFE can be rewritten as

$$Y'(D) = [X'(D) + Z(D)] \bmod [-\Delta/2, \Delta/2]. \qquad (16)$$

In order to analyze the performance limits of the equivalent channel, we lower bound the mutual information $I(X'; Y')$ between its input and output. The mutual information corresponding to a certain distribution on the channel's input

gives the highest possible rate for reliable communication, when a random channel code which is drawn according to that distribution is used [23]. However, the equivalent channel (16) is attained using a linear cyclic code (as opposed to a random code), and thus the mutual information may never be achieved. Nevertheless, the mutual information is a useful metric for the performance of ZF-IF if we account for the possible loss of rate due to the use of a linear cyclic code.

The channel (16) is a modulo-additive *colored* Gaussian noise channel, where the variance of the noise is $\sigma^2_{\text{ZF-IF-DFE}}$. The mutual information between the input and output of such a channel is maximized by a uniform memoryless distribution over the modulo interval. Moreover, for such an input distribution, the mutual information is lower bounded by that of a modulo-additive channel with *white* Gaussian noise having the same variance, see [24]. Let $\tilde{Z}$ be a Gaussian random variable with zero mean and variance $\sigma^2_{\text{ZF-IF-DFE}}$. If $X'$ has a memoryless uniform distribution over $[-\Delta/2, \Delta/2)$, which corresponds to taking $q$ to infinity, $Y'$ has a memoryless uniform distribution over the interval as well, and we have

$$
\begin{aligned}
I(X'; Y') &\geq \log_2(\Delta) - h(\tilde{Z} \bmod [-\Delta/2, \Delta/2)) \\
&\geq \log_2(\Delta) - h(\tilde{Z}) \qquad (17) \\
&= \frac{1}{2} \log_2(\Delta^2) - \frac{1}{2} \log_2(2\pi e \cdot \sigma^2_{\text{ZF-IF-DFE}}) \\
&= \frac{1}{2} \log_2 \left( \frac{\text{SNR}}{\sigma^2_{\text{ZF-IF-DFE}}} \cdot \frac{12}{2\pi e} \right),
\end{aligned}
$$

where (17) follows from the fact that modulo reduction can only decrease differential entropy. Let us express $\sigma^2_{\text{ZF-IF-DFE}}$ as a product of two factors

$$\sigma^2_{\text{ZF-IF-DFE}} = \gamma_{I,H} \cdot \sigma^2_{\text{ZF-DFE}},$$

where $\gamma_{I,H}$ is the additional noise enhancement caused by the FFE in ZF-IF w.r.t. that caused by the FFE of ZF-DFE. With this notation we have

$$
\begin{aligned}
I(X'; Y') &\geq \frac{1}{2} \log_2 \left( \frac{\text{SNR}}{\sigma^2_{\text{ZF-DFE}}} \cdot \frac{1}{\gamma_{I,H}} \cdot \frac{12}{2\pi e} \right) \\
&= \frac{1}{2} \log_2 \left( \frac{\text{SNR}/\Gamma_{I,H}}{\sigma^2_{\text{ZF-DFE}}} \right),
\end{aligned}
$$

where

$$\Gamma_{I,H} = \frac{2\pi e}{12} \cdot \gamma_{I,H}$$

is the SNR loss due to IF-ZF equalization. It follows that the gap-to-capacity in dB is given by[8]

$$10\log_{10}(\Gamma_{I,H}) = 10\log_{10}\left(\frac{2\pi e}{12}\right) + 10\log_{10}(\gamma_{I,H}). \quad (18)$$

The first term on the r.h.s. of (18) is the well-known high-SNR shaping gain, which equals 1.53dB. The reason for this loss is the one-dimensional modulo operation we use at the receiver. The modulo operation allows the decoder to use the original codebook $\mathcal{C}$ for decoding $\mathbf{x}'$ which preserves the original decoding complexity of the codebook. However, decoding the equivalent channel's output after the modulo

---

[7]The codebook $\mathcal{C}$ is invariant to the convolution with $I(D)$ as defined in Section III, and hence $X'(D)$ is uniformly distributed over the constellation (15), as is $X(D)$.

[8]This is the gap-to-capacity in the case of (complex) transmission over complex channels as well.

reduction is equivalent to searching for the point that was most likely transmitted over the *infinite* lattice $\mathcal{C} + \Delta\mathbb{Z}^N$. This is strictly suboptimal since no more than $2^{NR}$ points of the infinite lattice, which correspond to $\mathcal{C} \otimes \mathbf{i}$ (with operations carried over the *reals*), are valid, and hence better performance can be achieved by searching only over these valid points at the decoder.

While the modulo loss amounts to 1.53dB at high-SNR, at low SNR the loss may be significantly larger (see, e.g., [24]), which makes IF equalization less attractive in that regime. This loss can be mitigated by incorporating shaping at the transmitter, and/or avoiding the modulo reduction at the receiver. Implementing such modifications at the transmitter and receiver without significantly increasing the computational complexity is an interesting avenue for future research.

The second loss in (18) is related to the additional noise enhancement caused by the FFE. It follows from Theorem 1, that for the optimal choice of the integer-valued filter, $\gamma_H$ is upper bounded by

$$\gamma_H = \min_{I(D)} \gamma_{I,H} \le \min_{n \ge p+1}\left[\eta(n) \cdot (\alpha_H)^{\frac{1}{n}}\right]. \qquad (19)$$

In order to gain more insight regarding the term $\gamma_{I,H}$ we illustrate the effect of the noise enhancement through the following examples of ISI channels.

**Example 1 - Two-tap real "RAKE" channel:** An interesting example is a *real* channel with only two non-zero taps (which may be arbitrarily far apart), namely $H(D) = 1 + aD^p$. Without loss of generality, we can assume that $|a| \le 1$, as otherwise we can transform the channel into the channel $\tilde{H}(D) = 1 + \frac{1}{a}D^p$ by using an all-pass filter. It is rather obvious that the optimal choice for $I(D)$ is either $I_0(D) = 1$ or $I_1(D) = 1 + \frac{a}{|a|}D^p$. It can be shown by straightforward algebra that for the choice $I_0(D)$, the noise enhancement is given by $1/(1 - a^2)$, and for the choice $I_1(D)$ the noise enhancement is given by $2/(1 + |a|)$. It follows that for $|a| \le 1/2$, $I_0(D)$ is better, while for $1/2 < |a| \le 1$, $I_1(D)$ is better, and the maximum noise enhancement (which occurs for $|a| = 1/2$) is $4/3 \approx 1.25$dB. Since for this channel $\sigma^2_{\text{ZF-DFE}} = 1$, $\gamma_{I,H}$ is equal to the noise enhancement. Figure 4 depicts $\gamma_{I,H}$ for the real two-tap channel with an optimal choice of $I(D)$, along with the bound (19) for the same channel, and the noise enhancement caused by a ZF-LE. It is evident from the figure that for this channel the bound is not tight.

**Example 2 - Two-tap complex "RAKE" channel:** Consider the channel from the previous example $H(D) = 1 + aD^p$, where now $a = re^{j\theta}$ is a complex number. We assume without loss of generality that $|a| \le 1$. In contrast to the real-valued two-tap channel, in this case there is no clear choice of $I(D)$ for each value of $a$. Nevertheless, similar to the real two-tap channel, $\gamma_H$ is independent of the delay $p$ between the first and the second tap (as is also the case for the ZF-LE). This follows since the noise enhancement caused by each of the filters $A_1(D) = I(D)/H(D)$ and $A_2(D) = I(D^p)/H(D^p)$ is the same. Therefore, if for the channel $H(D) = 1 + aD$ a certain choice of $I(D)$ results in noise enhancement of $\gamma_{I,H}$,
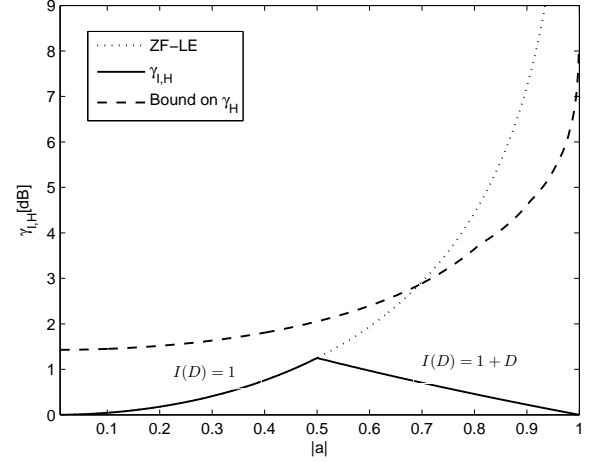


Fig. 4. $\gamma_{I,H}$ for the real two-tap channel $H(D) = 1 + aD^p$. The solid line shows $\gamma_{I,H}$ for the optimal choice of $I(D)$ for *any* $p$, the dashed line shows the bound (19) for $p=1$ and the dotted line shows the noise enhancement induced by ZF-LE

for the channel $\tilde{H}(D) = 1 + aD^p$, the choice $\tilde{I}(D) = I(D^p)$ results in $\gamma_{I,H}$ as well. While the performance of a ZF-LE is independent of the phase $\theta$, the noise enhancement caused by the ZF-IF equalizer is significantly influenced by $\theta$. It is interesting to note, however, that for the two-tap channel, the bound (19) (in its complex form) is independent of $\theta$. Figure 5 depicts $\gamma_{I,H}$ (in dB) for the two-tap channel with an optimal choice of $I(D)$ (which was found numerically), for values of $|a| \le 0.99$.
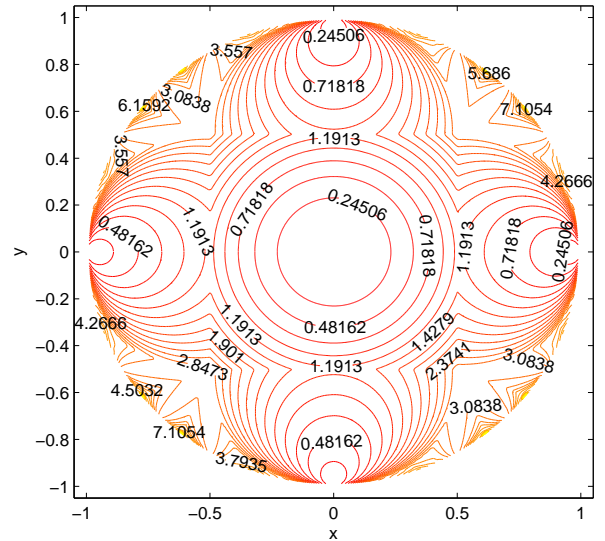


Fig. 5. $\gamma_{I,H}$ (in dB) for the complex two-tap channel $H(D) = 1 + (x + j \cdot y)D^p$, when the optimal integer-valued filter $I(D)$ is used.

**Example 3 - An ISI channel with random taps:** In this

example we consider the channel

$$H(D) = \sum_{k=0}^{p} h_k D^p,$$

where $\{h_k\}_{k=0}^{p}$ are real i.i.d. random Gaussian variables with zero mean and unit variance. We numerically evaluate the probability density function (pdf) of $\gamma_{I,H}$ in dB for the optimal choice of $I(D)$. Figure 6 depicts the results for $p = 3$, $p = 5$ and $p = 7$. The results show that $\gamma_H$ becomes larger when the channel is longer, which is also the behavior of the bound (19).
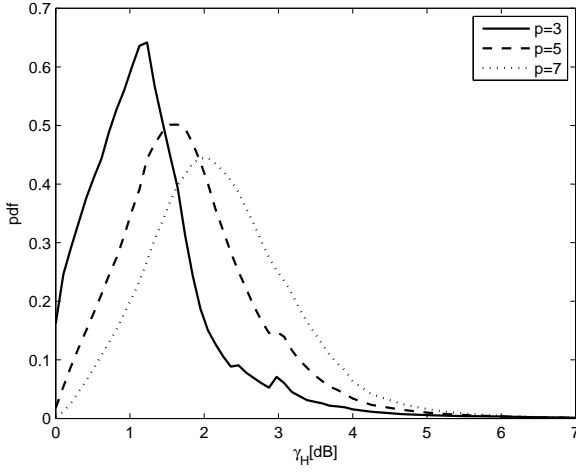


Fig. 6. The pdf of $\gamma_{I,H}$ (in dB) for the ISI channel $H(D) = \sum_{k=0}^{p} h_k D^p$ where $\{h_k\}_{k=0}^{p}$ are i.i.d. random Gaussian variables with zero mean and unit variance, when the optimal integer-valued filter $I(D)$ is used.

A third loss incurred by ZF-IF equalization is related to the constraint that the channel code is cyclic. The gap-to-capacity of ZF-IF equalization at a certain block error probability, is the sum of the shaping loss, the additional noise enhancement $\gamma_{I,H}$ (in dB) and the gap-to-capacity (in dB) of the chosen cyclic code at the desired block error probability. To the best of the authors knowledge, it is still an open question whether cyclic codes can attain the capacity of the AWGN channel, let alone, that of the mod-$\Delta$ AWGN channel. Nevertheless, algebraic cyclic codes (such as BCH for example) have been extensively used in communication systems for decades due to the fair tradeoff they offer between performance and complexity. Modern (e.g., LDPC) cyclic codes are a subject of extensive research nowadays, and some families of such codes were reported to have very good performances over the AWGN channel[9], see e.g. [25], [26].

Most of the research effort on linear cyclic codes have been devoted to binary codes. In the high-SNR regime, which is the focus of this paper, the cardinality of the code must be greater than two, and a binary code does not suffice. In the next section we propose a practical coded modulation scheme

[9] Note that in ZF-IF the additive noise after the FFE is not white, and thus it is important that the cyclic code should not be sensitive to the memory of the noise.

for high transmission rates that utilizes binary cyclic codes and is suitable for cyclic-coded integer forcing equalization.

## VII. PRACTICAL CODING AT HIGH RATES

In order to achieve high transmission rates, using a binary channel codebook, we transmit a combination of coded and uncoded bits which are mapped to a $2^M$-PAM constellation points by *natural labeling*. Namely, we transmit,[10]

$$\mathbf{x} = \mathbf{x}_c + \sum_{b=1}^{M-1} \mathbf{x}_{u_b} 2^b, \qquad (20)$$

where $\mathbf{x}_c$ is a codeword of length $N$ from the binary linear cyclic code $\mathcal{C}$, and $\mathbf{x}_{u_b}$, $b = 1, \ldots, M-1$, are blocks of uncoded information bits having the same length ($N$) as that of $\mathbf{x}_c$. Our goal is to detect $(\mathbf{x} \otimes \mathbf{i}) \bmod 2^M$ without error, and apply the DFE (with all operations carried over $\mathbb{Z}_{2^M}$) in order to reconstruct $\mathbf{x}$.[11]

We now show that natural labeling retains the closure property under integer-valued cyclic convolution.

*Lemma 3:* Let $\Lambda$ denote the set of vectors $\mathbf{x}$ given by (20) such that $\mathbf{x}_c \in \mathcal{C}$. Specifically,

$$\Lambda = \left\{ \mathbf{x} \mid \mathbf{x}_c \in \mathcal{C}, \mathbf{x}_{u_b} \in \mathbb{Z}_2^N, \ b = 1, \ldots, M-1 \right\}.$$

If $\mathbf{x} \in \Lambda$, then $(\mathbf{x} \otimes \mathbf{i}) \bmod 2^M \in \Lambda$.

*Proof:* Using natural labeling, it is clear that the result of the cyclic convolution reduced modulo $2^M$ can be written as,

$$(\mathbf{x} \otimes \mathbf{i}) \bmod 2^M = \mathbf{x}_0 + \sum_{b=1}^{M-1} \mathbf{x}_b 2^b,$$

for some vectors $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{M-1} \in \mathbb{Z}_2^N$. Clearly if $\mathbf{x}_0 \in \mathcal{C}$ then $(\mathbf{x} \otimes \mathbf{i}) \bmod 2^M \in \Lambda$. We have,

$$\begin{aligned} \mathbf{x}_0 &= \left( (\mathbf{x} \otimes \mathbf{i}) \bmod 2^M \right) \bmod 2 \\ &= (\mathbf{x} \otimes \mathbf{i}) \bmod 2 \\ &= \left( \left( \mathbf{x}_c + \sum_{b=1}^{M-1} \mathbf{x}_{u_b} 2^b \right) \otimes \mathbf{i} \right) \bmod 2 \\ &= \left( \mathbf{x}_c \otimes \mathbf{i} + \sum_{b=1}^{M-1} 2^b \left( \mathbf{x}_{u_b} \otimes \mathbf{i} \right) \right) \bmod 2 \\ &\overset{(a)}{=} (\mathbf{x}_c \otimes \mathbf{i}) \bmod 2 \overset{(b)}{\in} \mathcal{C}, \end{aligned}$$

where $(a)$ holds since $(\mathbf{x}_{u_b} \otimes \mathbf{i}) \in \mathbb{Z}^N$, and $(b)$ holds because $\mathcal{C}$ is a binary linear cyclic codebook. Thus, $\mathbf{x}_0$ is indeed a codeword. ∎

At the equalizer's output we get a corrupted version of $\mathbf{x} \otimes \mathbf{i}$, and after reducing modulo 2, we get $\mathbf{x}_0$ (plus folded Gaussian noise). Since $\mathbf{x}_0 \in \mathcal{C}$, it can be decoded before preceding to detect the uncoded bits. A complete knowledge of $\mathbf{x}_0$ divides the constellation into two cosets, which makes the distinction between two points from the same coset easier than the

[10] Due to the power constraint, the transmitted signal, in effect, would be given by (15) with $q = 2^M$.

[11] Note that in order to reconstruct $\mathbf{x}$ we need to know the value of its last $L$ samples. We therefore need to zero-pad the last $L$ uncoded bits as well as the last $L$ coded bits.

distinction between two points of the full constellation, thus doubling the Euclidean minimum distance in the constellation as in Ungerboeck's set partitioning (see, e.g., [1]). We can therefore first decode $\mathbf{x}_0$ and then detect the uncoded bits using a slicer with double step size. For more details, see [27], [28].

## VIII. DISCUSSION AND CONCLUSIONS

We have presented a novel DFE scheme for the discrete-time linear Gaussian channel, suitable for single-carrier transmission where channel state information is not available at the transmitter. The scheme enables block decoding to be performed before applying the DFE when a cyclic code is used. The channel is equalized to an impulse response that is comprised of integer coefficients only. The performance of the proposed scheme was analyzed, and in particular an upper bound on the induced noise enhancement was derived. Several examples of ISI channels were examined and it was shown that in many scenarios the scheme achieves a rather small gap-to-capacity. An interesting avenue for further research is to incorporate shaping into the scheme in order to make the scheme attractive at low SNR. Another interesting question to be explored, is how IF equalization can be combined with iterative equalization methods such as turbo equalization. Since the noise enhancement the FFE in IF equalization causes is always smaller (or equal) to that of an ZF-LE, combining it with an iterative equalizer may result in better convergence.

## ACKNOWLEDGEMENT

## APPENDIX

### THE MMSE-IF SCHEME

It is well known that the MMSE-DFE is strictly better than the ZF-DFE. The differences between the two are more pronounced at low SNR, and vanish as the SNR increases (for channels that satisfy the Paley-Wiener condition). We now develop an MMSE version for IF as well, which is strictly better than ZF-IF. Nevertheless, at high SNR the schemes have similar performance.

Our goal in designing the MMSE-IF equalizer is to produce an output,

$$x_k * i_k + e_k = \tilde{x}_k + e_k,$$

that maximizes the SINR, which we define as:

$$\text{SINR}_{\text{MMSE-IF-DFE}} = \frac{\mathbf{E}\left[\tilde{x}_k^2\right]}{\mathbf{E}\left[e_k^2\right]}.$$

Note that in fact we are interested in estimating $x'_k = x_k \otimes i_k$. However, since we transform the linear convolution performed by the channel into a cyclic one, simply by zero padding the transmitted block and ignoring the last entries of the output of the FFE (as described in Section III), there is no difference between estimating $\tilde{x}_k$ and $x'_k$.

The output of the FFE, $v_k$, serves as an estimate for $\tilde{x}_k$ from the sequence $\{y_k\}$, and can be written as,

$$
\begin{aligned}
V(D) &= A(D)Y(D) \\
&= A(D)H(D)X(D) + A(D)N(D) \\
&= I(D)X(D) + \left(\frac{A(D)H(D)}{I(D)} - 1\right)I(D)X(D) + Z(D) \\
&= \tilde{X}(D) + (G(D) - 1)\tilde{X}(D) + Z(D),
\end{aligned}
$$

where $Z(D)$ is the filtered Gaussian noise, and $G(D)$ is some linear time-invariant filter. Since we want our estimator to be unbiased (see [6]), we require that $G(D)$ be monic (i.e, $g_0 = 1$). In the time domain we have,

$$
\begin{aligned}
v_k &= \tilde{x}_k + \sum_{l \neq 0} g_l \tilde{x}_{k-l} + z_k \\
&= \tilde{x}_k + e_k. \tag{21}
\end{aligned}
$$

The derivation of the optimal MMSE-IF filter is reminiscent of finding the optimal MMSE-FFE in the classic DFE scheme [6]. However, there is one difference. In the classic DFE scheme the objective is to estimate the data symbols which are uncorrelated. Thus, the estimated symbol $x_k$ and the residual noise $e_k$ (which is composed of data symbols from different time instances and Gaussian noise) are also uncorrelated. In the MMSE-IF scheme, we estimate an (integral) *linear combination* of the data symbols. For this reason, different samples of $\tilde{x}_k$ are correlated and therefore $\tilde{x}_k$ and $e_k$ are correlated as well. We would like to be able to follow the derivation of the optimal MMSE-FFE [6]. To this end, we may remove the correlation between $\tilde{x}_k$ and $e_k$ by dithering as we show in Appendix A.

As shown in [6], if the output of an equalizer is composed of signal and noise which are uncorrelated, maximization of the SINR is equivalent to minimization of the MSE. This implies that the optimal choice for the front-end linear filter is given by the Wiener filter,

$$A(D) = b_0 \frac{I(D)H^*(D^{-*})}{H(D)H(D^{-*}) + \frac{1}{\sigma_x^2}}. \tag{22}$$

The role of the factor $b_0$ in (22) is to guarantee that the estimator is unbiased, i.e., to guarantee that $g_0 = 1$. Let $S_{ee}(e^{j\omega})$ be the Power Spectral Density (PSD) of the residual noise (ignoring the bias-removal term $b_0$). The filter $I(D)$ should be chosen such that,

$$
\begin{aligned}
\mathbf{E}\left[e_k^2\right] &= \frac{1}{2\pi}\int_{-\pi}^{\pi} S_{ee}(e^{j\omega})d\omega \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{|I(e^{j\omega})|^2}{|H(e^{jw})|^2 + \frac{1}{\sigma_x^2}}d\omega
\end{aligned}
$$

is minimized. Following the derivation of Section IV, we define,

$$
\tilde{K}_{\text{MMSE},n} = \begin{bmatrix}
\tilde{k}_0 & \tilde{k}_{-1} & \tilde{k}_{-2} \dots & \tilde{k}_{-(n-1)} \\
\tilde{k}_1 & \tilde{k}_0 & \tilde{k}_{-1} \dots & \tilde{k}_{-(n-2)} \\
\vdots & \vdots & \vdots \ddots & \\
\tilde{k}_{n-1} & \tilde{k}_{n-2} & \tilde{k}_{n-3} \dots & \tilde{k}_0
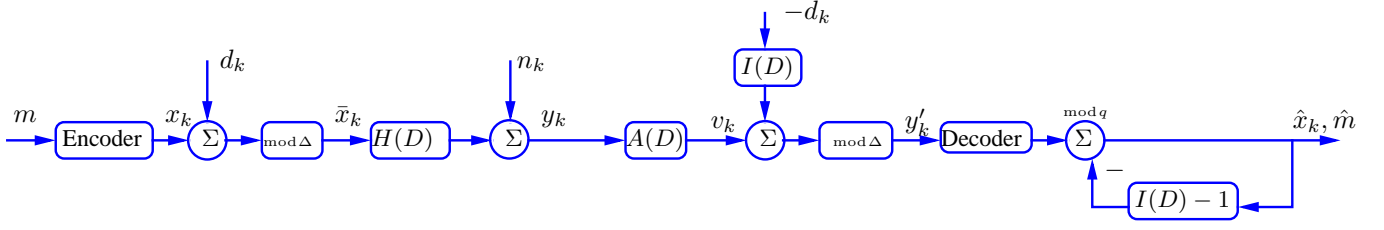\end{bmatrix},
$$

Fig. 7.   A schematic description of a MMSE-IF system. The notation $\mathrm{mod}\,\Delta$ denotes reducing modulo the interval $[-\Delta/2, \Delta/2)$

where

$$\tilde{k}_m = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{|H(e^{j\omega})|^2 + \frac{1}{\sigma_x^2}} e^{-jm\omega} d\omega$$

is the autocorrelation sequence corresponding to,

$$\tilde{K}(D) = \frac{1}{H(D)H^*(D^{-*}) + \frac{1}{\sigma_x^2}}.$$

Let $\tilde{\mathbf{F}}_{\mathrm{MMSE},n}$ be a matrix satisfying,

$$\tilde{\mathbf{K}}_{\mathrm{MMSE},n} = \tilde{\mathbf{F}}_{\mathrm{MMSE},n}^T \tilde{\mathbf{F}}_{\mathrm{MMSE},n}.$$

The optimal choice of $I(D)$ is given by finding the shortest vector in the lattice $\tilde{\mathbf{F}}_{\mathrm{MMSE},n}$, which can in practice be approximated by using the LLL algorithm as was explained in Section IV. Note that the choice of $I(D)$ is now SNR dependant.

### A. Removing Correlation by Dithering

In order to remove the correlation between $\tilde{x}_k$ and $e_k$ in (21), we use a pseudo-random dither $d_k$ which is uniformly distributed over the modulo interval $[-\Delta/2, \Delta/2)$, and is assumed to be known to both the transmitter and the receiver. As we recall, in the case of $q$-PAM transmission, i.e., the data symbols are

$$x_k \in c\sqrt{\mathrm{SNR}} \left\{ \frac{-(q-1)}{2q}, \frac{-(q-3)}{2q}, \ldots, \frac{q-1}{2q} \right\},$$

the modulo interval is simply $\left[ -c\sqrt{\mathrm{SNR}}/2, +c\sqrt{\mathrm{SNR}}/2 \right)$. The transmitted sequence is

$$\bar{x}_k = (x_k + d_k) \bmod [-\Delta/2, \Delta/2),$$

which is uniformly distributed over $[-\Delta/2, \Delta/2)$, and statistically independent of $x_k$. Nevertheless, we still seek an estimator for $\tilde{x}_k = x_k * i_k$.

Define $\tilde{\bar{x}}_k$ as,

$$\tilde{\bar{x}}_k = \bar{x}_k * i_k.$$

The receiver's front-end filter is the optimal linear MMSE estimator for $\tilde{\bar{x}}_k$ from the sequence $\{y_k\}$, given by,

$$A(D) = \tilde{b}_0 \frac{I(D)H^*(D^{-*})}{H(D)H(D^{-*}) + \frac{1}{\sigma_{\bar{x}}^2}},$$

where $\tilde{b}_0$ is a bias-removal term, serving the same purpose as $b_0$ in (22). The receiver then computes,

$$y'_k = (v_k - d_k * i_k) \bmod [-\Delta/2, \Delta/2)$$

$$= \left( \tilde{\bar{x}}_k + \sum_{l \neq 0} g_l \tilde{\bar{x}}_{k-l} + z_k - d_k * i_k \right) \bmod [-\Delta/2, \Delta/2)$$

$$= \Big( (x_k * i_k) \bmod [-\Delta/2, \Delta/2)$$

$$+ (d_k * i_k) \bmod [-\Delta/2, \Delta/2) - d_k * i_k$$

$$+ \sum_{l \neq 0} g_l \tilde{\bar{x}}_{k-l} + z_k \Big) \bmod [-\Delta/2, \Delta/2)$$

$$= \left( \tilde{x}_k + \sum_{l \neq 0} g_l \tilde{\bar{x}}_{k-l} + z_k \right) \bmod [-\Delta/2, \Delta/2)$$

$$= (\tilde{x}_k + e_k) \bmod [-\Delta/2, \Delta/2)$$

where $\tilde{x}_k$ and $e_k$ are uncorrelated. The complete system is illustrated in Figure 7.

### REFERENCES

[1] G. D. Forney, Jr. and G. Ungerboeck, "Modulation and coding for linear Gaussian channels," *IEEE Trans. Information Theory*, vol. IT-44, no. 7, pp. 2384–2415, Oct. 1998.

[2] E. A. Lee and D. G. Messerschmitt, *Digital Communication*. Boston, MA: Kluwer Academic, 1994.

[3] S. Litsyn, *Peak Power Control in Multicarrier Communications*. Cambridge: Cambridge University Press, 2007.

[4] S. H. Han and J. H. Lee, "An overview of peak-to-average power ratio reduction techniques for multicarrier transmission," *IEEE Wireless Commun.*, vol. 12, no. 2, pp. 56–65, Apr. 2005.

[5] T. Guess and M. K. Varanasi, "An information-theoretic framework for deriving canonical decision-feedback receivers in Gaussian channels," *IEEE Trans. Information Theory*, vol. IT-51, pp. 173–187, Jan. 2005.

[6] J. M. Cioffi, G. P. Dudevoir, M. V. Eyuboglu, and G. D. Forney, Jr., "MMSE decision-feedback equalizers and coding – Part I: Equalization results, Part II: Coding results," *IEEE Trans. Communications*, vol. 43, pp. 2582–2604, Oct. 1995.

[7] G. D. Forney, Jr., "Shannon meets Wiener II: On MMSE estimation in successive decoding schemes," in *Proc. 42nd Annual Allerton Conference on Communication, Control, and Computing*, Allerton House, Monticello, Illinois, Oct. 2004.

[8] M. Tomlinson, "New automatic equalizer employing modulo arithmetic," *Electron. Lett.*, vol. 7, pp. 138–139, Mar. 1971.

[9] J. Zhan, B. Nazer, U. Erez, and M. Gastpar, "Integer-forcing linear receivers," in *Proc. ISIT 2010*, Austin, Texas, June 2010, pp. 1022–1026.

[10] D. Yellia, A. Vardy, and O. Amrani, "Joint equalization and coding for intersymbol interference channels," *IEEE Trans. Information Theory*, vol. 43, no. 2, pp. 409–425, Mar. 1997.

[11] A. M. Chan and G. W. Wornell, "A class of block-iterative equalizers for intersymbol interference channels: fixed channel results," *IEEE Trans. Communications*, vol. 49, no. 11, pp. 1966–1976, Nov. 2001.

[12] S. Ariyavisitakul and Y. Li, "Joint coding and decision feedback equalization for broadband wireless channels," *IEEE Select. Areas Communications*, vol. 16, pp. 1670–1678, Dec. 1998.

[13] M. V. Eyuboglu, "Detection of coded modulation signals on linear, severely distorted channels using decision-feedback noise prediction with interleaving," *IEEE Trans. Communications*, vol. 36, pp. 401–409, Apr. 1988.

[14] K. Zhou, J. G. Proakis, and F. Ling, "Decision-feedback equalization of time-dispersive channels with coded modulation," *IEEE Trans. Communications*, vol. 38, pp. 14–24, Apr. 1990.

[15] P. Kabal and S. Pasupathy, "Partial-response signaling," *IEEE Trans. Communications*, vol. COM-23, pp. 921–934, Sep. 1975.

[16] R. F. H. Fischer and C. Siegl, "On the relation between lattice-reduction-aided equalization and partial-response signaling," in *Proc. International Zurich Seminar on Communications*, Zurich, Switzerland, Feb. 2006, pp. 34–37.

[17] E. Zehavi and J. K. Wolf, "On saving decoder states for some trellis codes and partial response channels," *IEEE Trans. Communications*, vol. 36, pp. 22–224, Feb. 1988.

[18] O. Shalvi, N. Sommer, and M. Feder, "Signal codes," *IEEE Trans. Information Theory*, Submitted June 2008, see http://arxiv.org/abs/0806.4773.

[19] A. K. Lenstra, H. W. Lenstra, and L. Lovasz, "Factoring polynomials with rational coefficients," *Math. Annalen*, vol. 261, pp. 515–534, 1982.

[20] D. Micciancio and S. Goldwasser, *Complexity of Lattice Problems: A Cryptographic Perspective*. Cambridge, U.K.: Kluwer Academic Publishers, Mar. 2002, vol. 671 of The Kluwer International Series in Engineering and Computer Science.

[21] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*. New York: Springer-Verlag, 1988.

[22] U. Grenander and G. Szego, *Toeplitz Forms and Their Applications*. Berkeley and Los Angeles: University of Callifornia Press, 1958.

[23] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 1991.

[24] U. Erez and S. T. Brink, "A close-to-capacity dirty paper coding scheme," *IEEE Trans. Information Theory*, vol. 51, no. 10, pp. 3417–3432, Oct. 2005.

[25] Y. Kou, S. Lin, and M. P. C. Fossorier, "Low-density parity-check codes based on finite geometries: A rediscovery and new results," *IEEE Trans. Information Theory*, vol. IT-47, pp. 2711–2736, Nov. 2001.

[26] C. Tjhai, M. Tomlinson, M. Ambroze, and M. Ahmed, "Cyclotomic idempotent-based binary cyclic codes," *Electron. Lett.*, vol. 41, no. 6, pp. 341–343, Mar. 2005.

[27] O. Ordentlich and U. Erez, "Achieving the gains promised by integer-forcing equalization with binary codes," in *Proc. 26th Annual Convention of Electrical and Electronic Engineers in Israel*, Eilat, Israel, Nov. 2010, pp. 703–707.

[28] O. Ordentlich, J. Zhan, U. Erez, M. Gastpar, and B. Nazer, "Practical code design for Compute-and-Forward," in *Proc. ISIT 2011*, Saint Petersburg, Russia, Aug. 2011.

[29] M. Tuchler, R. Koetter, and A. Singer, "Turbo equalization: Principles and new results," *IEEE Trans. Communications*, vol. 50, pp. 754–767, May 2002.

[30] H. Yao and G. W. Wornell, "Lattice-reduction-aided detectors for MIMO communication systems," in *Proc. IEEE Globecom 2002*, Taipei, Taiwan, Nov. 2002.